Mortality control charts for comparing performance of surgical units: validation study using hospital mortality data

Paris P Tekkis, Peter McCulloch, Adrian C Steger, Irving S Benjamin, Jan D Poloniecki

Abstract

Objective To design and validate a statistical method for evaluating the performance of surgical units that adjusts for case volume and case mix.
Design Validation study using routinely collected data on in-hospital mortality.
Data sources Two UK databases, the ASCOT prospective database and the risk scoring collaborative (RISC) database, covering 1042 patients undergoing surgery in 29 hospitals for gastro-oesophageal cancer between 1995 and 2000.
Statistical analysis A two level hierarchical logistic regression model was used to adjust each unit's operative mortality for case mix. Crude or adjusted operative mortality was plotted on mortality control charts (a graphical representation of surgical performance) as a function of number of operations. Control limits defined as 90%, 95%, and 99% confidence intervals identified units whose performance diverged significantly from the mean.
Results The mean in-hospital mortality was 12% (range 0% to 50%). The case volume of the units ranged from one to 55 cases a year. When crude figures were plotted on the mortality control chart, four units lay outside the 90% control limit, including two outside the 95% limit. When operative mortality was adjusted for risk, three units lay outside the 90% limit and one outside the 95% limit. The model fitted the data well and had adequate discrimination (area under the receiver operating characteristics curve 0.78).
Conclusions The mortality control chart is an accurate, risk adjusted means of identifying units whose surgical performance, in terms of operative mortality, diverges significantly from the population mean. It gives an early warning of divergent performance. It could be adapted to monitor performance across various specialties.

Introduction

Public concern in the United Kingdom after the Bristol inquiry into cardiac surgery is reflected in mounting pressure for open scrutiny of surgical outcomes.[1] For some major types of surgery, operative mortality is an important measure of performance. To reflect performance accurately, however, mortality must be adjusted for the effect of pre-existing comorbid disease. Existing models of risk stratification have several problems. Increasing specialisation of surgery means that regression models developed from "general surgical" cohorts are inappropriate. Existing models are also poor at interpreting large fluctuations in crude mortality caused by a few deaths in units with a small volume of surgery. Lastly, the assumption that relations between predictive variables and mortality are identical across units may obscure factors affecting mortality that are specific to particular units.

Gastrectomy and oesophagectomy have the highest mortality among elective operations in Britain. Patients with gastro-oesophageal cancer often have other serious conditions that increase the risks of surgery. The provision of surgery for upper gastrointestinal cancer is undergoing major reorganisation in Britain, favouring subspecialisation and centralisation and causing major changes in the case mix of surgery units. Directly comparing operative mortality in specialist units with a high volume of elective surgery with mortality in district hospitals with a low volume of high risk gastrointestinal emergencies can be misleading. Evidence about the relation between case volume and outcome conflicts.[2-4] The subspecialty of upper gastrointestinal cancer surgery exemplifies the general problem of quantifying surgical risk with adjustment for case mix and volume. We developed statistical techniques for evaluating surgical performance on a continuous scale and applied the techniques to data on upper gastrointestinal cancer surgery.

Data and methods

Data sources

We took data on outcomes of gastro-oesophageal cancer surgery from two databases on upper gastrointestinal surgery: the stomach and oesophageal cancer outcome and techniques (ASCOT) prospective database and the risk scoring collaborative (RISC) database. There was no population overlap between the databases. Both databases provided comprehensive POSSUM (physiological and operative severity score for the enumeration of mortality and morbidity) data on large cohorts of gastro-oesophageal surgery patients.[5]

Academic Department of Surgery, King's College Hospital, London SE5 9RS
Paris P Tekkis *research fellow of the Royal College of Surgeons of England*
Irving S Benjamin *professor of surgery*

Academic Unit of Surgery, University of Liverpool, University Hospital Aintree, Liverpool L9 7AL
Peter McCulloch *senior lecturer in surgery*

Department of Surgery, University Hospital Lewisham, London SE13 6LH
Adrian C Steger *consultant surgeon*

Department of Public Health Sciences, St George's Hospital, London SW17 0QT
Jan D Poloniecki *senior lecturer in biostatistics*

Correspondence to:
P P Tekkis
ptekkis@blueyonder.co.uk

bmj.com 2003;326:786

The ASCOT prospective database—This database on gastro-oesophageal cancer surgery, which was developed by the British Oesophago-Gastric Cancer Group, collects a comprehensive dataset on cases of gastro-oesophageal cancer referred to surgeons, whether or not an operation actually took place.[6] The data include patients' demographic details, preoperative assessment, tumour staging, type of surgery, postoperative course, and pathology. For this study the database's coordinator used an independent source (hospital episode statistics) to validate a sample of 157 cases. From January 1999 to December 2000 the 31 hospitals across the United Kingdom that joined this voluntary collaboration submitted data on 1036 cases.

The RISC database—This database recorded data on 601 patients undergoing oesophageal and gastric surgery in five hospitals in the South East and Thames Region, which included cases from general and thoracic surgical units. Of the cases, 351 were recorded retrospectively from pre-existing databases, case notes, theatre books, and operating lists, and 250 were prospectively collected from January 1999 to January 2001. The data were independently validated against other hospital data sources (medical records or mortuary registers).

Inclusion and exclusion criteria

We included data on oesophageal and gastric operations for malignant and benign disease with palliative or curative intent. We excluded cases where patients were treated medically or by endoscopic techniques (n=572) and cases with missing notes (n=23).

End point and risk factors

The primary end point was in-hospital mortality (any death during the same hospital admission as the operation), which can be more reliably quantified than 30 day mortality and includes patients with complications who remained in hospital beyond 30 days. Risk factors studied were age; sex; POSSUM score; surgical procedure (as classified by the Office of Population Censuses and Surveys' list of surgical operations and procedures, fourth revision (OPCS4)[7]; mode of surgery (emergency or elective); tumour staging (according to the International Union Against Cancer (UICC) system, fifth edition)[8]; and malignancy (according to POSSUM category).

Statistical analysis

We used univariate analysis to identify risk factors for mortality. Continuous variables were grouped into subcategories, and unifactorial logistic regression was used to compare these with a reference level. We used the χ^2 test to analyse categorical variables. To maximise information extracted by the model, we used the multiple imputation technique to substitute for incomplete data.[9 10]

We used a multifactorial logistic regression model to adjust for different hospitals' case mix. We constructed a two level hierarchical regression model to allow for clustering of outcomes among patients from the same hospital. Risk factors, including their interaction terms relating to individual patients, were entered into the first level of the model, while hospitals constituted the second level of the model, whose coefficients were allowed to vary randomly between units. We calculated expected mortality for each unit by excluding each unit in turn and modelling the remaining centres (a cross validatory approach).[11] The ratio of observed to expected mortality for each unit was multiplied by the mean mortality from the pooled data to derive each unit's risk adjusted operative mortality. We used a non-parametric bootstrap resampling technique with 10 000 iterations to calculate standard errors and to correct parameter estimation bias. We calculated exact binomial 95% confidence intervals for the observed mortality and risk adjusted operative mortality for each unit.

Validation of the model—To evaluate the performance of the model we used the Hosmer-Lemeshow $\hat{c}$ statistic to assess calibration or goodness of fit (the ability of the model to assign correct outcome probabilities to individual patients) and the area under the receiver operating characteristics (ROC) curve to assess discrimination (the ability of the model to assign higher risks to patients who die than to patients who live).[12 13] Values for the area under the ROC curve from 0.7 to 0.8 indicate reasonable discrimination and values exceeding 0.8 indicate good discrimination.

Mortality control chart—This graphical method for monitoring surgical performance plots units' mortality as a function of number of operations. The exact binomial distribution is used to construct control limits (90%, 95%, and 99% confidence intervals) around the mean operative mortality for the group. These control limits indicate whether a particular unit's operative mortality differs significantly from the mean at 10%, 5%, and 1% significance levels. Each unit's operative mortality (unadjusted or adjusted for case mix) can be plotted as a single point representing the total mortality or as a running mean as a function of the number of operations done. Underperforming units will lie above the upper control limits, while units with unusually good results will lie below the lower control limits. Units lying within the 95% control limits have an operative mortality that is statistically consistent with the group mean.

Statistical software—We used Intercooled STATA 6.0 for Windows (StataCorp, College Station, TX), NORM Version 2.03 for Windows (Pennsylvania State University, PA), and MLwiN Version 2.1c (University of London, London).

Results

Of 1637 cases, 1042 (63.7%) satisfied the inclusion criteria: 497 of 1036 cases (47.9%) in the ASCOT database and 545 of 601 cases (90.7%) in the RISC database. Although 36 hospitals contributed data to the study, the analysis was based on data from 29 centres, as seven units did not contribute operated cases and were therefore excluded. The cases comprised 538 oesophagectomies (51.6%), 443 gastrectomies (42.5%), and 61 palliative bypass procedures (5.9%) (table 1). Of the operations, 828 (79.5%) were elective and 78 (8.6%) were emergencies; in 136 cases (13.1%) the mode of surgery was not recorded. Nine hundred and nineteen operations (93.7%) were for cancer. The overall in-hospital operative mortality was 12% (9.4% in patients having an elective procedure and 26.9% in patients having an emergency procedure). No evidence of systematic under-reporting of risk factors was

shown, and missing data were distributed evenly among the hospitals.

We used the two level hierarchical logistic model, together with the overall median regression line, to calculate the relations between age of patients and operative mortality (figure 1) and between preoperative POSSUM physiological score and operative mortality for each of the 29 hospitals (figure 2). Case mix (based on POSSUM scores) varied significantly across units, as shown in figure 2 by the different ranges in POSSUM score (Kruskal-Wallis test: χ^2=62.159, df=28, P<0.0001).

The final multifactorial model used age, POSSUM score, POSSUM malignancy category, and mode of surgery as risk factors (table 2). Mode of surgery was retained in the model as it is clinically highly relevant and has been reported as an important predictor of outcome.[2] The model fitted the data well (Hosmer-Lemeshow ĉ statistic: χ^2=0.139, df=8, P=0.255) and had adequate discrimination (area under the ROC curve 0.78 (standard error 0.02)).

Units reported between one and 55 operations a year, with mortality ranging from 0% to 50%. The mortality control chart for unadjusted operative mortality shows that four units lay outside the 90% control limit (figure 3). When operative mortality was adjusted for case mix, however, no unit was shown to underperform at the 95% control limit, and the individual values regress towards the mean (figure 4). Two units had better results than the group average, with risk adjusted operational mortalities of 4.2% and 3.8%. Figure 5 shows the running means of the risk adjusted operational mortality for two of the units (31 and 33), representing two consecutive series of 102 and 166 cases. Despite fluctuations, unit 31 remained within the central part of the graph, whereas unit 33 repeatedly crossed the lower 99% control limit and thus could be said to be a truly outlying unit and a consistently good performer.

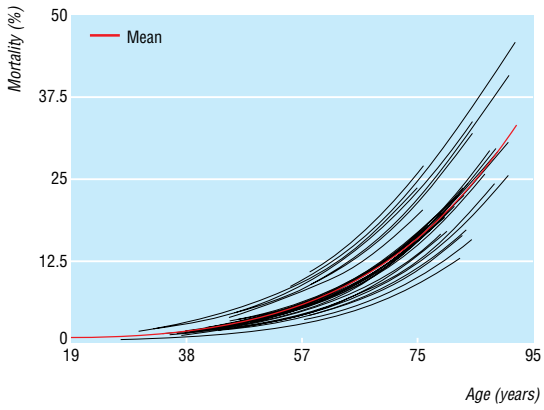

Fig 1 Relation between operative mortality and age of patients, shown as individual prediction curves for the 29 hospitals

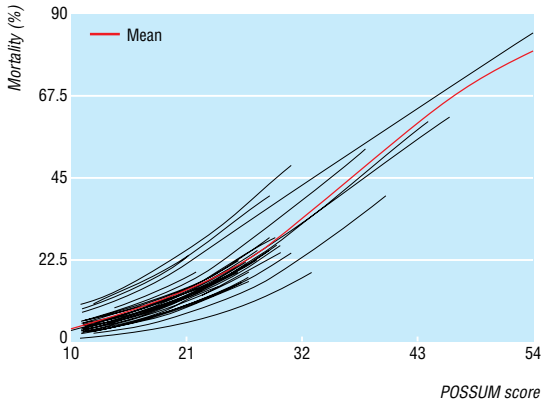

Fig 2 Relation between operative mortality and patients' preoperative POSSUM score, shown as individual prediction curves for the 29 hospitals

Table 1 Mortality among patients undergoing upper gastrointestinal cancer surgery

Risk factor	No of patients (% of total)	No of deaths (% mortality)	Unadjusted odds ratio (95% CI)*
Age (years)†			
<60	284 (27.3)	13 (4.6)	1
61-70	294 (28.2)	29 (9.9)	2.28 (1.16 to 4.48)
71-80	351 (33.7)	61 (17.4)	4.38 (2.36 to 8.16)
>80	74 (7.1)	17 (23)	6.22 (2.86 to 13.52)
Data missing	39 (3.7)	5 (12.8)	
Sex			
Female	297 (28.5)	31 (10.4)	1
Male	688 (66)	88 (12.8)	1.28 (0.82 to 1.94)
Data missing	57 (5.5)	6 (10.5)	
POSSUM score†			
11-14	410 (39.3)	27 (6.6)	1
15-20	360 (34.5)	43 (12)	1.92 (1.16 to 3.18)
21-30	203 (19.5)	41 (20.2)	3.59 (2.12 to 6.03)
>30	24 (2.3)	10 (41.7)	10.1 (4.12 to 24.93)
Data missing	45 (4.3)	4 (8.9)	
Tumour staging			
I	126 (12.1)	14 (11.1)	1
IIa/b	294 (31.6)	24 (8.2)	0.71 (0.36 to 1.43)
IIIa	350 (37.6)	36 (10.3)	0.92 (0.48 to 1.77)
IIIb	33 (3.5)	7 (21.2)	2.15 (0.8 to 5.87)
IV	116 (12.5)	24 (20.7)	2.08 (1.02 to 4.26)
No tumour or data missing	123 (11.8)	20 (16.3)	
POSSUM malignancy category			
Primary only	355 (34.1)	31 (8.7)	1
Nodal disease	460 (44.1)	55 (12)	1.42 (0.89 to 2.26)
Metastatic disease	104 (10)	21 (20.2)	2.64 (1.44 to 4.84)
No malignancy	62 (6)	10 (16.1)	2.01 (0.93 to 4.34)
Data missing	61 (5.9)	8 (13.1)	
Mode of surgery			
Elective	828 (79.5)	78 (9.4)	1
Emergency	78 (7.5)	21 (26.9)	3.542 (2.04 to 6.15)
Data missing	136 (13.1)	26 (19.1)	
Type of surgery			
Oesophagectomy:	538 (51.6)	46 (8.6)	1
Right two stage	297 (28.5)	20 (6.7)	
Thoracoabdominal	106 (10.2)	9 (8.5)	
McKeown three stage	22 (2.1)	4 (18.1)	
Transhiatal	45 (4.3)	6 (13.3)	
Other	68 (6.5)	7 (10.3)	
Gastrectomy:	443 (42.5)	68 (15.3)	1.94 (1.30 to 2.89)
Total	162 (15.5)	35 (21.6)	
Subtotal	230 (22.1)	25 (10.9)	
Completion	3 (0.3)	1 (33.3)	
Wedge resection	2 (0.2)	1 (50)	
Other	47 (4.5)	5 (10.6)	
Gastrojejunostomy	61 (5.9)	11 (18)	2.35 (1.15 to 4.83)
Total	1042	125 (12)	

*Odds ratios and 95% confidence intervals were calculated by unifactorial logistic regression analysis. For the three major types of surgery (oesophagectomy, gastrectomy, and palliative gastrojejunostomy) univariate analysis was used.
†For purposes of illustration the continuous variables of age and POSSUM score were grouped into meaningful subgroups.

Table 2 Two level hierarchical logistic regression model for upper gastrointestinal surgery in all 29 hospitals

Risk factor	Coefficient β	Standard error	Odds ratio (95% CI)*
Age (per 10 year increase)†	0.48	0.11	1.62 (1.28 to 1.99)
POSSUM score†	0.77	0.17	2.16 (1.54 to 3.03)
POSSUM malignancy category‡:			
Primary only	0.27	0.37	1.35 (0.60 to 2.56)
Nodal disease	0.53	0.42	1.78 (0.70 to 3.64)
Metastatic disease	1.04	0.50	3.14 (1.06 to 7.16)
Mode of surgery‡:			
Emergency	0.57	0.37	1.78 (0.88 to 3.82)
Constant	–7.24	0.96	—
Level 2 variance	0.26	0.16	—

*Bias corrected bootstrap confidence intervals (10 000 iterations).
†Odds ratios and 95% confidence intervals are shown for every 10 year increase in age or 10 unit increase in POSSUM score.
‡Reference categories: no malignancy for POSSUM malignancy category; elective surgery for mode of surgery.

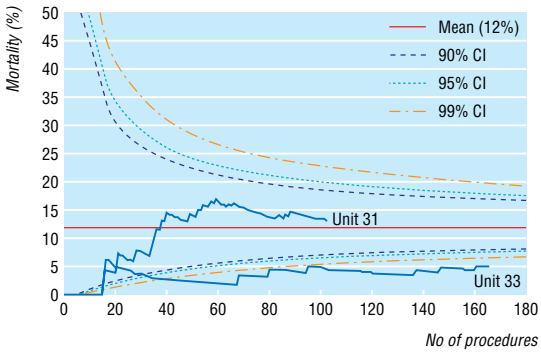

Fig 5 Operative mortality in units 31 (n=102) and 33 (n=166), plotted as running means (adjusted for case mix)

Discussion

The mortality control chart improves on current methods of evaluating surgical units' performance. It is an accurate, risk adjusted means of identifying outlying units while giving an early warning of units approaching divergence from the mean.

Validity of the data

The information in the study was a combination of prospective data and medical records. Centres voluntarily contributed data, and at present there is no formal system for externally validating the completeness of the database. Internal validity was established by comparing the operative mortality for a random sample of five participating hospitals (157 patients) with hospital episode statistics obtained independently from the hospitals' information departments. The two databases reported similar overall mortality (14% in the ASCOT data and 13.8% in the hospital episode statistics), but they differed in the individual hospitals' volumes of operations and in the variability of mortality. Although overall operative mortality in the units in our study was consistent with recently published data from the West Midlands region, our units were not randomly selected, and we cannot be sure how representative they are of all UK hospitals. However, although the quality of our data is limited, implementation of such a monitoring system in hospitals should lead to an increased awareness of the data that need to be collected, with subsequent improvement in the quality of the data.

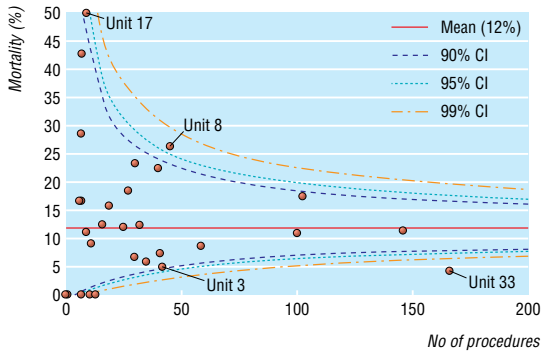

Fig 3 Operative mortality in 29 hospitals, unadjusted for case mix

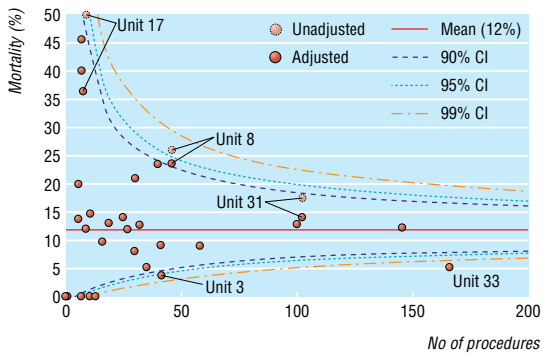

Fig 4 Operative mortality in 29 hospitals, adjusted for case mix (with unadjusted mortality for three hospitals shown)

Quality of the statistical analysis

Hierarchical regression models are particularly useful in modelling observations with a hierarchical or clustered structure, such as patients in different hospitals or pupils in different schools.[14] Such models avoid the penalty for ignoring the clustered nature of data on patients in hospitals—namely, an erroneously low standard error of regression coefficients.[15] Hierarchical models acknowledge heterogeneity among units and assume that the variability between hospitals approximates a normal distribution.[16] Such techniques have been adopted to rank the performance of organisations.[1 17 18] We used confidence intervals around the providers' performances to compare each unit's performance with the average, with wider confidence intervals for low volume units. If these wider limits are not allowed for, low volume providers are more likely to be ranked misleadingly at the top or bottom of the group. Confidence intervals can be placed around a unit's rank, thus emphasising "the caution with which any league tables must be treated."[19]

Control limits in the mortality chart define outlying units and give an early warning when a unit's performance starts to diverge from the population mean. Mortality charts can express either performance over a period as a point estimate or sequential monitoring, using running means of operative mortality. Crossing the upper control limit indicates high mortality and crossing the lower control limit indicates low mortality that is not attributable to normal variation. In each case efforts should be made to identify the special causes.

What is already known on this topic

League tables are an established technique for ranking the performance of organisations such as healthcare providers

Mortality control charts are another way to compare the performance of healthcare providers, particularly for outcomes of surgery

What this study adds

Mortality control charts can be adjusted for case mix and case volume and are better than league tables for monitoring surgical performance

Mortality control charts have a "buffer zone" for indicating divergence from the mean mortality and are particularly useful for specialties with a low volume of surgery

The less extreme control limits delineate an early warning "buffer zone" to trigger examination of practice. Because the control limits are much wider for low volumes, a high (risk adjusted) operational mortality in these hospitals should be interpreted carefully and may require longer monitoring to establish a meaningful estimate of mortality.

Usefulness of mortality control charts

Mortality control charts can be extended to any surgical specialty that uses risk adjusted outcomes. Similar graphical methods have been used to investigate the effect of case volume on unadjusted operative mortality in paediatric cardiac surgery.[20] Control charts based on the approach of Walter Shewhart—the pioneer of the economic control of variation in manufacturing—have been described for monitoring surgical performance.[21 22] Other studies have described alternative techniques based on the cumulative sum (CUSUM) technique for longitudinal analysis of surgical performance.[23 24] In the sequential mortality control chart (figure 5), type I errors will occur more often but can be reduced by using the more extreme control limits and by interpreting divergences with proper caution. The mean operative mortality and corresponding control limits for any population will need to be reviewed periodically to reflect changes over time. The mortality control chart is intended to add to existing statistical methods for monitoring surgical performance rather than replace them.

We thank all the consultants who contributed data to the study, the data collection officers for their help, and the research staff at the Centre for Multilevel Modelling, University of London, for their invaluable help in developing the hierarchical models. Hospitals and trusts that contributed data were Addenbrooke's NHS Trust, Aintree University Hospital, Airedale NHS Trust, Barnet General Hospital, Bishop Auckland General Hospital, Broomfield Hospital, Chorley General Hospital, Colchester General Hospital, Furness General Hospital, Glenfield Hospital, Harefield Hospital, Harrogate District Hospital, Huddersfield Royal Infirmary, Ipswich Hospital, Kingston Hospital, Leicester Royal Infirmary, Leighton Hospital, Macclesfield District General Hospital, Maidstone General Hospital, Newham General Hospital, Norfolk and Norwich NHS Trust, North Staffordshire City General Hospital, Queen Alexandra Hospital, Queen Elizabeth Hospital, Queen Mary's Hospital, Royal Bolton Hospital, Royal Bournemouth Hospital, Royal Free Hospital, Royal Hull Hospitals NHS Trust, Royal Lancaster Infirmary, University Hospital Lewisham, Watford General Hospital, West Wales General Hospital.

Contributors: PPT, ISB, and JDP devised the original research and obtained funding. PPT, ACS, and PMcC were responsible for the completion and validation of the RISC and ASCOT datasets respectively. PPT and JDP analysed the data. PPT drafted and edited the paper and JDP and PMcC revised it. All authors contributed comments and corrections on the final draft. PPT and PMcC are the guarantors for the study.

Funding: The Hue Falwasser Fellowship of the Royal College of Surgeons of England. The guarantors accept full responsibility for the conduct of the study, had access to the data, and controlled the decision to publish.

Competing interests: None declared.

Ethical approval: The multicentre research ethics committee for Wales.

1 Spiegelhalter DJ, Aylin P, Best NG, Evans SJW, Murray GD. Commissioned analysis of surgical performance by using routine data: lessons from Bristol inquiry. *J R Statist Soc (Ser A)* 2002;165:1-31.
2 Gillison EW, Powell J, McConkey CC, Spychal RT. Surgical workload and outcome after resection for carcinoma of the oesophagus and cardia. *Br J Surg* 2002;89:344-8.
3 Swisher SG, Deford L, Merriman KW, Walsh GL, Smythe R, Vaporicyan A, et al. Effect of operative volume on morbidity, mortality, and hospital use after esophagectomy for cancer. *J Thorac Cardiovasc Surg* 2000;119:1126-32.
4 Birkmemeyer JD, Siewers AE, Finlayson EVA, Stukel TA, Lucas FL, Batista I, et al. Hospital volume and surgical mortality in the United States. *N Engl J Med* 2002;346:1128-37.
5 Copeland GP, Jones D, Walters M. POSSUM: a scoring system for surgical audit. *Br J Surg* 1991;78:355-60.
6 Cummins J, McCulloch P. ASCOT: a comprehensive clinical database for gastro-oesophageal cancer surgery. *Eur J Surg Oncol* 2001;27:709-13.
7 Department of Health. *Hospital episode statistics: main operations 2000/01.* www.doh.gov.uk/hes/ (accessed 13 Dec 2002).
8 Sobin LH, Wittekind C. *TNM classification of malignant tumours.* 5th ed. New York: John Wiley & Sons, 1997.
9 Harrell FE, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med* 1996;15:361-87.
10 Schafer JL. *NORM: Multiple imputation of incomplete multivariate data under a normal model, version 2.* www.stat.psu.edu/~jls/misoftwa.html (accessed 13 Dec 2002).
11 Aylin P, Alves B, Best N, Cook A, Elliott P, Evans SJ, et al. Comparison of UK paediatric cardiac surgical performance by analysis of routinely collected data 1984-96: was Bristol an outlier? *Lancet* 2001;358:181-7.
12 Hosmer DW, Lemeshow S. *Applied logistic regression.* 2nd ed. New York: John Wiley & Sons, 2000.
13 Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 1982;143:29-36.
14 Goldstein H, Thomas S. Using examination results as indicators of school and college performance. *J R Statist Soc (Ser A)* 1996;159:149-63.
15 Greenland S. Principles of multilevel modelling. *Int J Epidemiol* 2000;29:158-67.
16 Marshall EC, Spiegelhalter DJ. Institutional performance. In: Leyland AH, Goldstein H, eds. *Multilevel modelling of health statistics.* Chichester, W Sussex: John Wiley & Sons, 2001:127-42.
17 Parry GJ, Gould CR, McCabe CJ, Tarnow-Mordi WO. Annual league tables of mortality in neonatal intensive care units: longitudinal study. International Neonatal Network and the Scottish Neonatal Consultants and Nurses Collaborative Study Group. *BMJ* 1998;316:1931-5.
18 Daley J, Khuri SF, Henderson W, Hur K, Gibbs JO, Barbour G, et al. Risk adjustment of the postoperative morbidity rate for the comparative assessment of the quality of surgical care: results of the National Veterans Affairs surgical risk study. *J Am Coll Surg* 1997;185:328-40.
19 Marshall EC, Spiegelhalter DJ. Reliability of league tables of in vitro fertilisation clinics: retrospective analysis of live birth rates. *BMJ* 1998;316:1701-4.
20 Stark J, Gallivan S, Lovegrove J, Hamilton JR, Monro JL, Pollock JC, et al. Mortality rates after surgery for congenital heart defects in children and surgeons' performance. *Lancet* 2000;355:1004-7.
21 Mohammed MA, Cheng KK, Rouse A, Marshall T. Bristol, Shipman, and clinical governance: Shewhart's forgotten lessons. *Lancet* 2001;357:463-7.
22 Adab P, Rouse AM, Mohammed MA, Marshall T. Performance league tables: the NHS deserves better. *BMJ* 2002;324:95-8.
23 Poloniecki J, Valencia O, Littlejohns P. Cumulative risk adjusted mortality chart for detecting changes in death rate: observational study of heart surgery. *BMJ* 1998;316:1697-700.
24 Lovegrove J, Valencia O, Treasure T, Sherlaw-Johnson C, Gallivan S. Monitoring the results of cardiac surgery by variable life-adjusted display. *Lancet* 1997;350:1128-30.

(Accepted 6 February 2002)